OIPE JC100 AUG 15 2002 PATENT & TRADEMARK OFFICE

MAIL DATE CANCELLED OIPE JC100 AUG 14 2002 PATENT & TRADEMARK OFFICE

RECEIVED
AUG 20 2002
Technology Center 2100

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Applicant: Agrawal	)	Art Unit: 2172
Serial No.: 09/487,191	)	Examiner: Fleurantin
Filed: January 19, 2000	)	**AM9-99-0226**
For: **SYSTEM AND ARCHITECTURE FOR PRIVACY-PRESERVING DATA MINING**	)	August 13, 2002 750 B STREET, Suite 3120 San Diego, CA 92101

APPEAL BRIEF

Commissioner of Patents and Trademarks
Washington, DC 20231

Dear Sir:

This appeal brief is submitted under 35 U.S.C. §134. This appeal is further to Appellant's Notice of Appeal filed herewith.

Table of Contents

Section	Title	Page
(1)	Real Party in Interest	1
(2)	Related Appeals/Interferences	2
(3)	Status of Claims	2
(4)	Status of Amendments	2
(5)	Summary of Invention	2
(6)	Issues	2
(7)	Grouping of Claims	3
(8)	Argument	3
App.A	Appealed Claims	

(1) Real Party in Interest

The real party in interest is IBM Corp.

08/28/2002 BHILLIAR 00000004 090441 09487191
01 FC:120 320.00 CH

(2) Related Appeals/Interferences

No other appeals or interferences exist which relate to the present application or appeal.

(3) Status of Claims

Claims 1-13 and 20-23 are pending and finally rejected as being obvious over Fayyad et al. in view of Tendick.

(4) Status of Amendments

No amendments are outstanding.

(5) Summary of Invention

As exemplified in Claim 1, the invention in one aspect is a computer-implemented method for obtaining data from at least one user computer via the Internet while maintaining the privacy of a user of the computer. The invention of Claim 1 includes perturbing original data associated with the user computer to render perturbed data, and using a distribution of the perturbed data, generating at least one estimate of a distribution of the original data. Using the estimate of the distribution of the original data, a data mining model is generated.

(6) Issue

Whether Claims 1-13 and 20-23 are unpatentable under 35 U.S.C. §103 as being obvious over Fayyad et al. in view of Tendick.

(7) Grouping of Claims

The claims stand together.

(8) Argument

Claims 1-13 and 20-23 have been rejected under 35 U.S.C. §103 as being obvious over Fayyad et al. in view of Tendick et al., the secondary reference having been used as a reference of the teaching of privacy.

At issue is whether Fayyad et al.'s teaching of perturbing a single value - a mean - for a different purpose - finding a starting point of a centroid - is the same as the limitation in Claim 1 of perturbing a distribution of original values to preserve user privacy while finding a model. The examiner, relying on the guidance of MPEP §2111, which promotes broad claim interpretation during prosecution, believes that it is.

The examiner's error is easy to see. MPEP §2111 cannot trump the other sections of the MPEP, and read properly, it does not. MPEP §2111 comes accompanied by a cautionary limit not mentioned in the Office Action, namely, that claim terms are to be construed **as a skilled artisan** (as opposed to an examiner intent on preserving a rejection) would construe them, MPEP §2111.01 (citing In re Sneed). The relied-upon section of the MPEP is not a license to assign words any meaning that are convenient for depriving a claim of patentability.

With the actual words of MPEP §2111 in mind, there has been no showing supported by actual evidence that a skilled artisan would construe a distribution of values (recited in the plural) to be the same thing as a mean of the values, In re Dembiczak, 175 F.3D 994, 50 U.S.P.Q.2d 1614 (Fed. Cir. 1999) (the range of sources available does not diminish the requirement for **actual evidence**, and "broad conclusory

statements regarding the teaching of multiple references, standing alone, are not evidence"). In fact, a single value representing a mean of many values is a different mathematical concept than the actual distribution of the values themselves, much less is the idea of perturbing a mean of otherwise unperturbed values the same thing as perturbing the distribution of the values.

Having deconstructed the prima facie case as currently fashioned by the examiner, Appellant will present a proper analysis of the claims vis-a-vis the cited references in accordance with the MPEP and the law. What is of more importance than the examiner's misguided understanding of MPEP §2111 is MPEP §2141 et seq., setting forth the law of obviousness in Graham v. John Deere. Specifically, MPEP §2143 requires that, to establish a prima facie case of obviousness, an examiner must

> (1) identify a suggestion to modify the cited reference(s) as proposed by the Examiner to arrive at the instant claimed invention;
>
> (2) set forth why the Examiner believes that a reasonable expectation of success exists for the proposed modification of the references which would be necessary to arrive at the claimed subject matter; and
>
> (3) explain where in the prior art all of the claimed limitations are taught or suggested.

Here, no prior art showing of a perturbation of a distribution of values has been shown, only the perturbation of the singular mean of unperturbed values, much less has any rational reason accompanied by the requisite prior art support been given why the skilled artisan would be motivated to import privacy-driven considerations into Fayyad et al. which nowhere considers the issue. Indeed, the rejection itself is internally inconsistent. On page 2, the examiner takes the position that Appellant is wrong that Fayyad et al. does not use perturbed original values, while on page 3 he states that it would have been obvious to change Fayyad et al. to use perturbed original values, implicitly admitting that Fayyad et al. does not have such a feature but must be modified to have it. These are two separate issues requiring divergent analyses, neither of which has been properly followed, but in any case Appellant is left wondering what really is the examiner's position? Does Fayyad et al. have the claimed perturbed distribution as asserted on page 2 or must it be modified to have it as contended on page 3? Hopefully, the examiner will clarify in his Answer, and Appellant will then duly respond in a Reply Brief.

Using the proper obviousness analysis set forth above, Claim 1 recites using a distribution of perturbed data to generate an estimate of a distribution of the original data, and using the estimate of the distribution of the original data to generate a data mining model, whereas Fayyad et al. is directed only to using a perturbed value of a ***mean of original values***, not the original values themselves, to find a starting centroid for candidate data clusters, col. 2, lines 4-5. Per Fayyad et al. this is "cheaper" than doing a full-blown cluster operation, col. 5, lines 61-65. Fayyad et al. does not use or suggest using perturbed values of original values at all. Instead, once the starting centroids are found, Fayyad et al. uses actual original data to generate the models. This is not surprising, because Fayyad et al. nowhere considers privacy, but is only concerned with finding a starting point for its centroids.

With this understanding of Fayyad et al. in mind, there has been no teaching cited for perturbing a distribution of values, much less in the context of Claim 1, as otherwise required by step (3) above. Further, step (1) of MPEP §2143 cannot be complied with using the art at hand because none of the art provides motivation to use a perturbed distribution of original values instead of using a single perturbation of a single mean for any reason. All Fayyad et al. seeks to do is to identify a starting point for its centroids, a concept requiring only a single datum. Why would the skilled artisan want to use an entire perturbed distribution for this purpose? Indeed, using an entire perturbed distribution just to find the starting point of a centroid would be useless, wasteful, and inadvisable, thus depriving the prima facie case of compliance with step (2). Quite so: finding a "cheaper" way to identify starting points of centroids, as taught by Fayyad et al., col. 5, lines 61-65, teaches away from substituting an entire perturbed distribution for a single mean datum. Since a reference must be considered for all it teaches, including teachings away (MPEP §2141.02, citing W.L. Gore & Assoc. v. Garlock), the examiner's position that it would have been obvious to modify Fayyad et al. to use an entire distribution instead of a single mean is legally wrong.

Respectfully submitted,

John L. Rogitz
Registration No. 33,549
Attorney of Record
750 B Street, Suite 3120
San Diego, CA 92101
Telephone: (619) 338-8075

JLR:jg

APPENDIX A

1. A computer-implemented method for obtaining data from at least one user computer via the Internet while maintaining the privacy of a user of the computer, comprising the acts of:

perturbing original data associated with the user computer to render perturbed data;

using a distribution of the perturbed data, generating at least one estimate of a distribution of the original data; and

using the estimate of the distribution of the original data, generating at least one data mining model.

2. The method of Claim 1, wherein perturbed data is generated from plural original data associated with respective plural user computers.

3. The method of Claim 2, wherein the original data cannot be reconstructed from the respective perturbed data.

4. The method of Claim 2, wherein at least some of the data is perturbed using a uniform probability distribution.

5. The method of Claim 2, wherein at least some of the data is perturbed using a Gaussian probability distribution.

6. The method of Claim 2, wherein at least some of the data is perturbed by selectively replacing the data with other values based on a probability.

7. A computer system including a program of instructions including structure to undertake method acts comprising:

at a user computer, randomizing at least some original values of at least some numeric attributes to render perturbed values;

sending the perturbed values to a server computer not having access to the original values; and

at the server computer, processing the perturbed values to generate at least one classification model.

8. The computer of Claim 7, wherein perturbed values are generated from plural original values associated with respective plural user computers.

9. The computer of Claim 7, wherein the original values cannot be reconstructed from the respective perturbed values.

10. The computer of Claim 7, wherein at least some of the original values are perturbed using a uniform probability distribution.

11. The computer of Claim 7, wherein at least some of the original values are perturbed using a Gaussian probability distribution.

12. The computer of Claim 7, wherein the method acts further comprise perturbing categorical values of at least some categorical attributes by selectively replacing the categorical values with other values based on a probability.

13. A computer storage device including computer readable code readable by a server computer for generating at least one classification model based on original data values stored at plural user computers without knowing the original values, comprising:

logic means for receiving perturbed values from the user computers, the perturbed values representing randomized versions of the original values; and

logic means for generating at least one classification model using at least in part the perturbed values and not using the original values.

20. The method of Claim 1, further comprising sending the model to at least one user computer for use thereof by the user computer on original data.

21. The method of Claim 20, wherein the user computer uses the model on original data to render a classification, and then sends the classification to the Web site.

22. The method of Claim 20, wherein the model is sent to the user computer as a JAVA applet.

23. The system of Claim 7, wherein the method acts further comprise:

sending the model to at least one user computer for use thereof by the user computer on original data.